

1586/BENEFITS AND CONSEQUENCES OF USING A SINGLE PORTUGUESE LANGUAGE VERSION OF THE FUNCTIONAL ASSESSMENT OF CANCER THERAPY-GENERAL (FACT-G)

Benjamin Arnold, S. Eremenco & J.S. Lai, CORE, ENH, Evanston, IL; J.L.P. Ribeiro, Faculdade de Psicologia, U. do Porto, Porto, Portugal; Maribel Dóro, BMT Unit, University Hospital, Curitiba, Paraná, Brazil; D. Cella, CORE, ENH, Evanston, IL

Aims: Languages such as Spanish and Portuguese are spoken differently in different parts of the world. It is challenging to determine whether to adapt language versions to account for regional differences in vocabulary or syntax, or whether one universal translation can be applied with due attention to harmonization of differences. To evaluate our practice of creating and applying single language translations in countries where different variations of the same language are spoken we compared the psychometric properties of the Portuguese FACT-G as completed by people from Portugal versus Brazil. **Methods:** Version 4 of the FACT-G was translated into one Portuguese version for both countries using our standard methodology. In Brazil, data were collected from 143 cancer patients, 51.1% male, mean age: 57.0 years. In Portugal, data were collected from 122 cancer patients, 37.7% male, mean age: 57.2 years. We analyzed the data using classical test theory and then applied a rating scale item response theory model to examine differential item functioning (DIF). **Results:** The FACT-G performed equally well in both countries with Cronbach's $\alpha = 0.89$ in each sample, and all subscales between 0.71 and 0.86. Among the 27 items of the FACT-G, 10 items displayed DIF between the groups ($p < 0.01$). The majority of DIF was found in the social family and emotional well-being subscales as opposed to the physical and functional well-being subscales. **Conclusions:** The Portuguese patients were more homogeneous than the Brazilian sample in terms of functioning and diagnosis, but this alone did not account for the divergences observed. The results suggest cultural (rather than linguistic) differences between groups that pertain to emotional and family norms. Although it is feasible to develop one Portuguese language version of the FACT-G for use in both Portugal and Brazil, more investigation is needed to appreciate cultural differences in the way Brazilian and Portuguese patients respond to psychosocial questions. Steps aimed at addressing these needs will be discussed.

1206/TEST-RETEST RELIABILITY OF HEALTH UTILITIES INDEX SCORES: EVIDENCE FROM HIP FRACTURE

David H. Feeny, Pharmacy and Pharmaceutical Sciences, Institute of Health Economics; Allyson Jones, Physical Therapy, University of Alberta; Ken Eng, Institute of Health Economics, Edmonton, AB, Canada

Aims: There is little published evidence on the test-retest reliability of utility scores derived from multi-attribute utility measures such as the Health Utilities Index. A prospective cohort study of patients recovering from hip fracture provided an opportunity to assess test-retest reliability. Patients experienced little change in the period between 3 and 6 months post hip fracture. The aim was to estimate test-retest reliability for overall HUI Mark 2 (HUI2) and Mark 3 (HUI3) utility scores. **Methods:** An inception cohort of hip fracture patients were recruited. Patients were enrolled within 3–5 days of surgery. Baseline assessments included the Functional Independence Measure (FIM[®]) and the HUI2/HUI3 questionnaire. Follow-up assessments at 1, 3, and 6 months also included a global change question. Two approaches were used to estimate test-retest reliability using the intra-class correlation (ICC) coefficient: by assessing agreement among scores in stable patients and with generalizability theory. Patients were classified as stable if their FIM overall scores changed by 10 points or less and if the patient classified themselves as having changed little between the 3 and 6 month assessments. Agreement for all patients at the 3 and 6-month assessments was used for generalizability theory approach. **Results:** Complete data at both the 3 and 6 months assessments were available for 195 patients; 141 patients were classified as stable. The ICCs for stable patients for HUI2 and HUI3 were 0.71 and 0.72; the ICCs derived from generalizability theory were 0.76 and 0.77. **Conclusions:** Test-retest reliability for HUI in this cohort was similar to reliability estimates for other preference-based multi-attribute and generic profile measures

of health status and health-related quality of life and in the acceptable range for making valid group-level comparisons.

1543/ITEM RESPONSE THEORY TO ASSESS THE FRANCE – SPAIN CROSS-CULTURAL VALIDITY OF THE VSP-A

Karine Barrau, Public Health, University Hospital of Marseille, Marseille, France; V. Serra-Sutton & L. Rajmil, Agència d'Avaluació de Tecnologia, Barcelona, Spain; P. Auquier, S. Robitail & M.C. Simeoni, Public Health, University Hospital of Marseille, Marseille, France

Aims: The VSP-A is a French generic, multidimensional, self-administered HRQoL tool, based on the point of view of adolescents. It comprises 34 items, exploring 9 dimensions: Friends (FR), psychological well-being (PWB), vitality (VI), self perception (PH), parents (FA), teachers (TE), school work (SCH), leisure (LE) and physical well-being (PH). VSP-A has already been validated in Spanish. The aim of this study is to explore the cross-cultural properties of this instrument by the means of IRT. **Methods:** The scalability of every dimensions was assessed using both non parametric and parametric IRT: Mokken scale analysis (MSP[®]) and Rasch model (WINSTEPS[®]). The multidimensional structure was assessed through Exploratory Factor Analyses (EFA), Confirmatory Factor Analyses (LISREL) and MAP analyses. DIF was assessed through non parametric Double Monotonicity Model and Zumbo's logistic regression method. Trait validity has been assessed using Multitrait Multimethod model (CFA). Known group validity has been checked comparing age, gender and different health status groups. The study involved 1356 adolescents (France $n = 783$, Spain $n = 573$). **Results:** LISREL CFA results were satisfactory (RMSEA = 0.074, CFI = 0.95). Non parametric IRT Loevinger H was greater than 0.40 for all of the dimensions on international and national data. MAP analyses and EFA showed good results. Partial Credit Rasch Model fit and DIF analyses were satisfactory for all of the items except one (INFIT > 1.40, Delta- $R^2 > 3\%$). Multitrait multimethod CFA test of fit was good (CFI > 0.90). Reliability is satisfactory on both the international sample (α : 0.71–0.91), and the national samples (France: 0.73–0.91; Spain: 0.68–0.92). Known group validity was achieved using several comparisons between age, gender groups and health status groups. **Conclusions:** These results confirm the cross-cultural validity of the Spanish and French versions of the VSP-A. The conjoint use of non parametric and parametric IRT model is useful and should be widely reported, especially when performing DIF analyses.